

single Data sets

Data Analysis 3







































general mathematics preliminary course

name: _____

PRELIMINARY CAPACITY MATRIX - GENERAL MATHEMATICS




























TOPIC: Data Analysis 3 - Single Data Sets

2 weeks

CONTENT	CAPACITY BREAKDOWN!	DONE IT!!!!	GOT IT!!!!	ON MY WAY!	WORKING ON IT!	HELP!!!!
1. Formative assessment	Watdoyaknow sheet					
2. Creation of tally charts and frequency tables to organise ungrouped and grouped data	L1: Ex 9A Q1-6					
	L2: Ex 9A Q2, 4, 6-11					
3. Creation of dot plots, sector graphs, bar graphs, column graphs and line graphs with emphasis on scaling of the axes.	L1: SS9.4 Ex 9B Q1, 3, 5, 7, 8, 11, 14					
	L2: SS9.4 Ex9B Q2,4,6-8,11, 12, 14, 15					
4. Misrepresentation of data						
5. Linking types of data with appropriate displays	Collating Yr 11 survey = Spreadsheet activity					
6. Radar charts	Ex 9B Q 9, 10, 13					
7. Frequency graphs and cumulative frequency graphs	L1: Ex 9C Q1, 2, 3, 6-16 SS9.5, Ex 9D Child Development W/S					
	L2: Ex 9C Q1a-3a, 4-17 SS9.5 Ex 9D Child Development W/S					
8. Division of data into deciles and quartiles						
9. Range and Interquartile range						
10. Stem and leaf plots to represent small data sets	Ex 9E					
11. Five number summary and box and whisper plots	Ex 9F					
12. Describing strengths/weaknesses of various graphs	Collating Yr 11 survey and collection of other graphs					

FORMATIVE MATRIX - GENERAL MATHEMATICS

TOPIC: Data Analysis 3 - Single Data Sets

QUESTION & ACCOMPANYING CONTENT	GOT IT!!!!	HAD SOME HELP!	HELP!!!!
1. Frequency distributions			
2. Frequency distributions			
3. Interpreting graphs			
4. Constructing bar charts			
5. Constructing sector graphs			
6. Interpreting graphs			
7. Constructing a dot plot			
8. Constructing frequency histograms and polygons			
9. Constructing cumulative frequency histograms and polygons (ogive)			

frequency tables

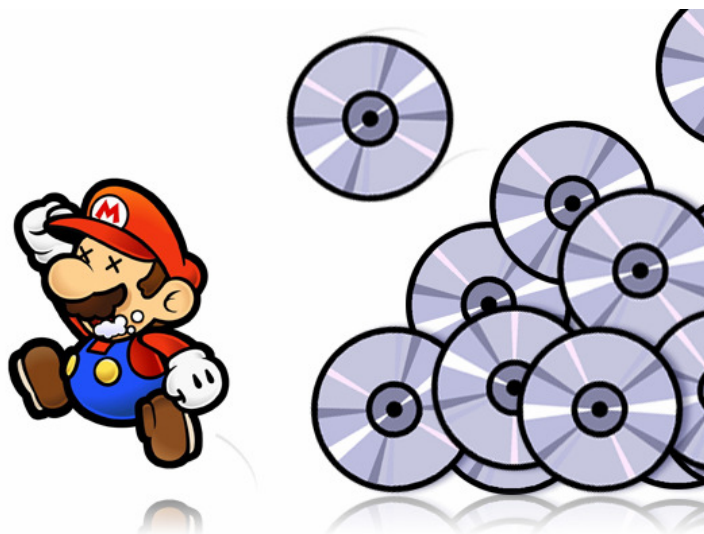
A **frequency table** is used to tabulate statistical data. A well-designed table allows the person doing the statistical investigation to easily tally up the data. In this topic we are concerned with the tallying of quantitative data.

eg Fifty people were surveyed and asked the number of DVDs that they had hired from a Video store in the past month. The results are shown below.

2 4 5 3 3 3 3 0 5 1
1 3 3 0 5 2 2 1 3 1
3 4 4 2 0 1 4 0 3 2
0 1 5 5 1 2 3 5 3 4
0 2 0 4 4 4 2 2 4 5

Enter the information in a frequency table.

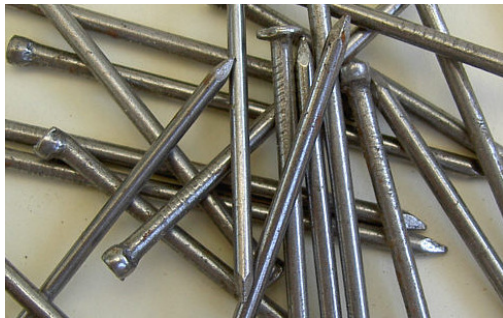
SCORE	TALLY	FREQUENCY



Sometimes it would be pointless trying to tally the data as individual (ungrouped) scores. It makes better sense to group the scores.

eg Consider the following data that give the number of nails in each of a sample of 40 boxes:

130	122	118	139	126	128	119	124	122	123
132	138	129	139	116	123	126	128	131	142
137	134	126	129	127	118	130	132	134	132
137	124	134	134	120	137	141	118	125	129



LOOK!

- We group the data in **CLASSES**.
- The class sizes **MUST BE THE SAME SIZE!!!!**
- An extra column should be added to show the **CLASS CENTRE**.
- The **CLASS CENTRE** is the middle of each class.

No. of nails	Class Centre	Tally	Frequency

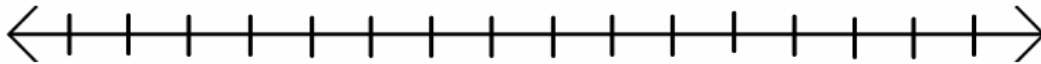
TYPES of graphs

Dot plots

eg Below are the scores out of 10 achieved by 15 students on a Maths quiz:

7 4 9 7 6 7 4 8 5 2 7 9 5 8 7

Display this information on a dot plot.



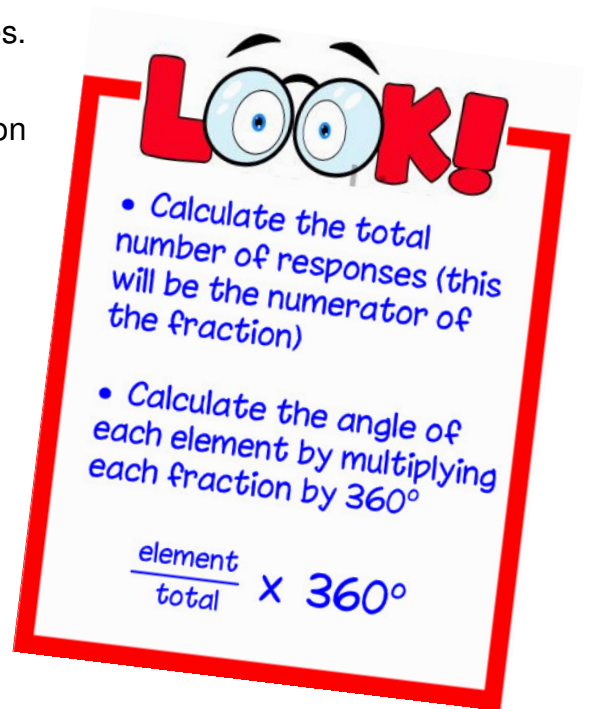
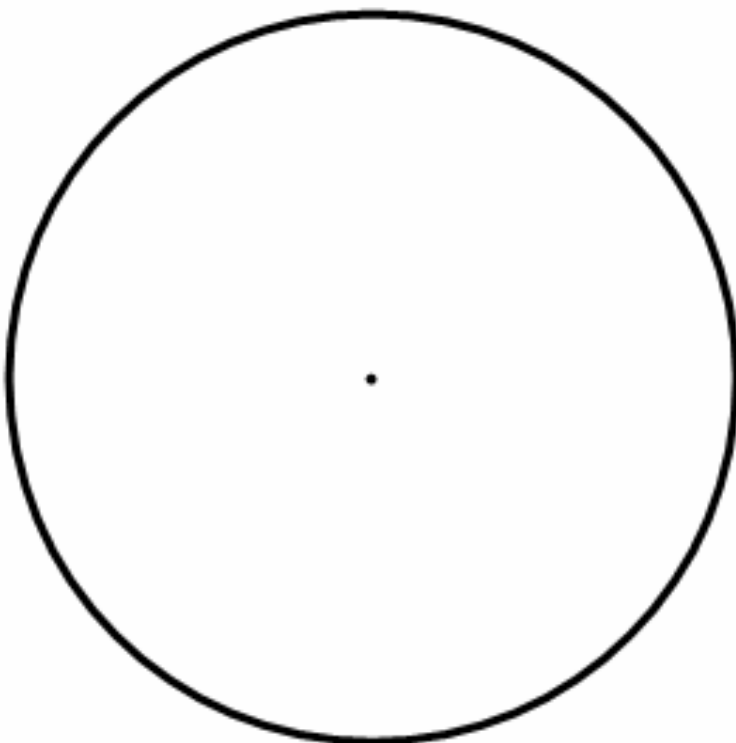
Sector graphs

 are used to compare quantities.

eg A poll was conducted on an issue of justice based on the statement "Apple has every right to track their phones and ipads". The responses were:

Agree 24 Disagree 57 Unsure 9

Present this information in a sector graph.

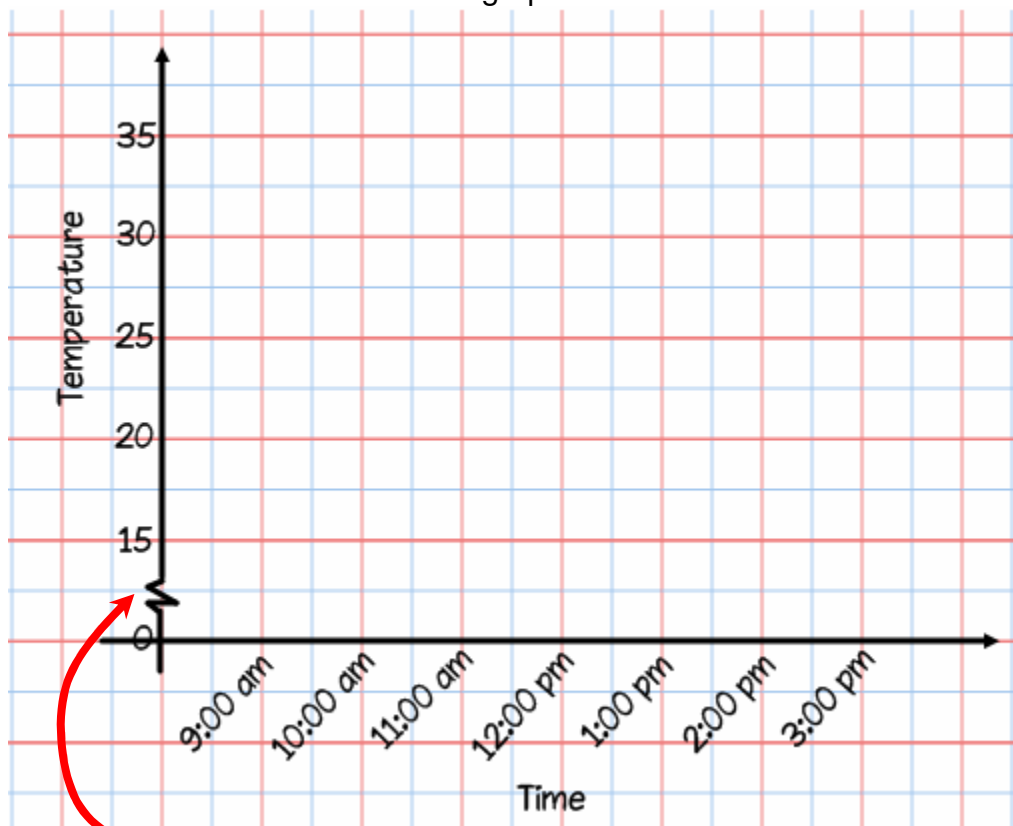


Line graphs are used to compare the change in one quantity with the change in another. It is suitable for **quantitative data**.

eg The table below shows the temperature taken in a classroom over a day.

Time	9:00am	10:00	11:00	Noon	1:00pm	2:00	3:00
Temp (C)	17	21	25	28	33	33	31

Present this information as a line graph.



Bar and Column graphs are used when we want to show a quantity and compare it to other quantities.

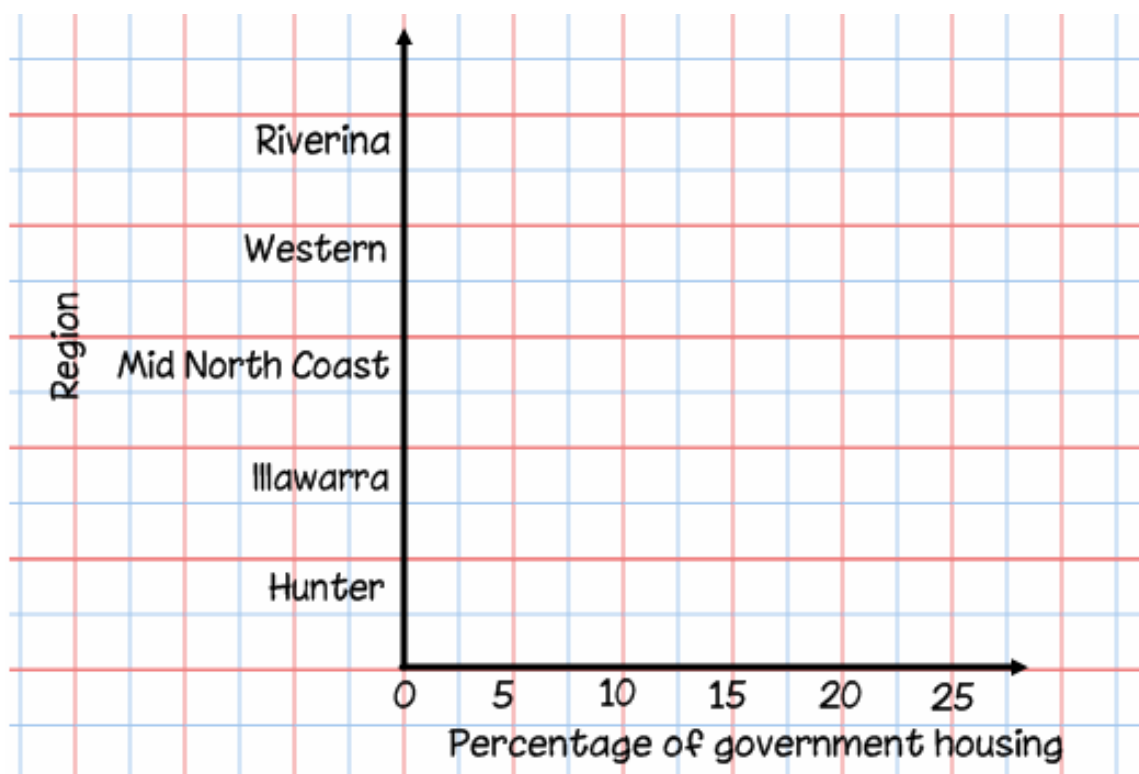
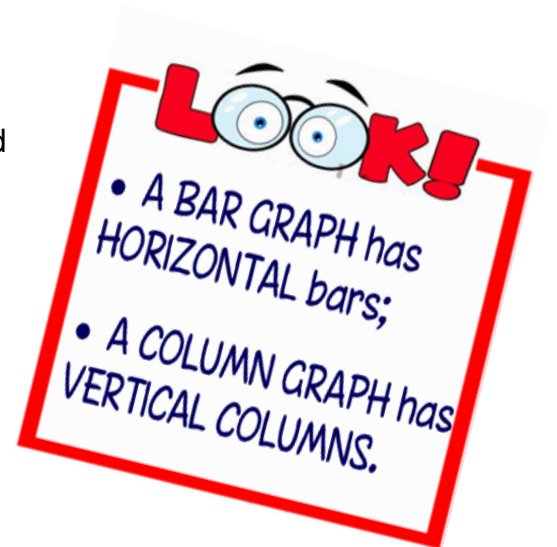
This is suitable for categorical data.

A **bar graph** has the categories on the **vertical axis** and frequency on the **horizontal axis**. These axes are **reversed** for a column graph.

eg The information shown below gives the government housing rates in various NSW regions:

Hunter	16.5%
Illawarra	23.1%
Mid North Coast	8.4%
western	10%
Riverina	11.6%

Show this information in a bar graph.



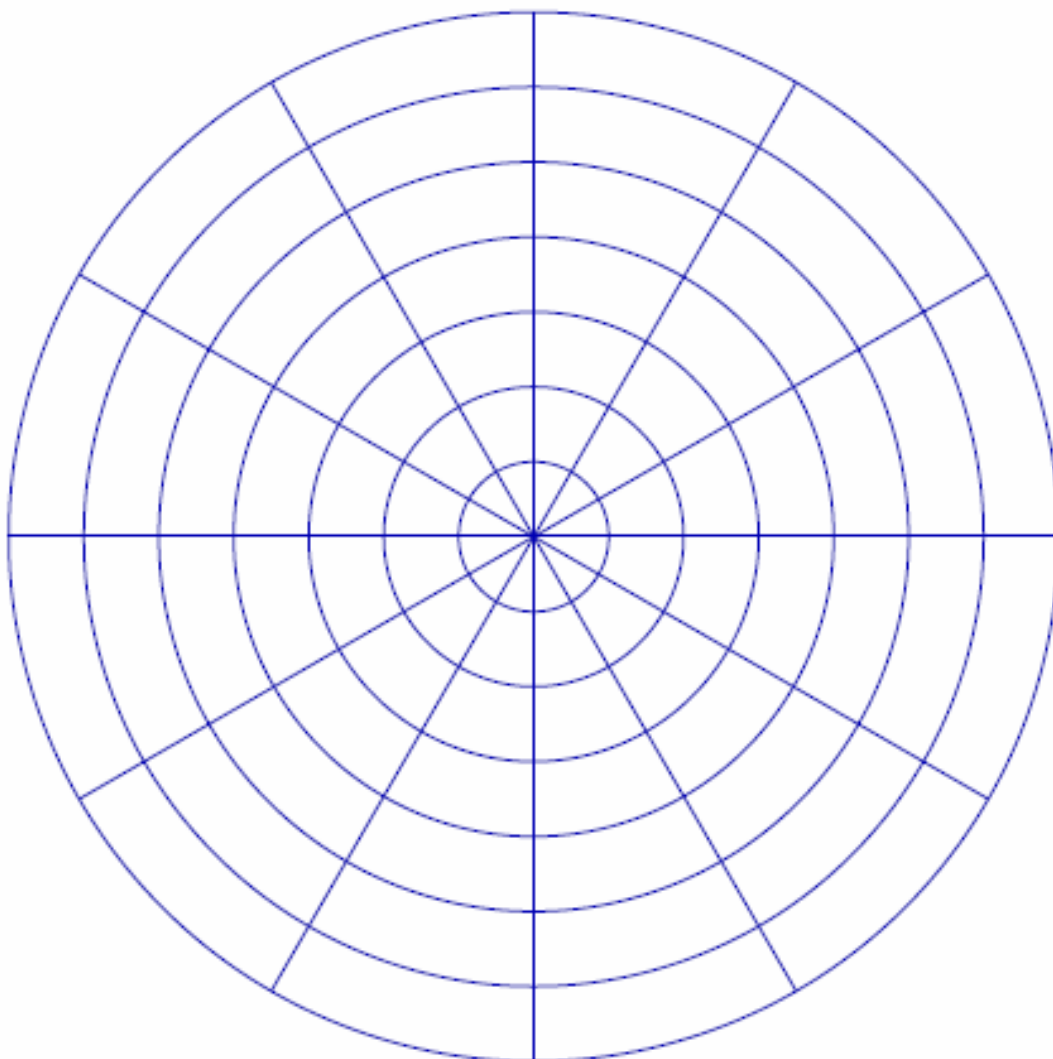
A Radar chart is similar to a line graph except it is circular. It is particularly suitable for presenting data trends that repeat.

The 'radar' is drawn with the data being measured placed in equal sectors around the circle and the results having a scale from the centre. The points are then plotted and joined consecutively.

eg The information below shows the sales in a department store over a year.

Month	Sales (\$m)	Month	Sales (\$m)
January	2.8	July	1.8
February	1.7	August	1.1
March	1.1	September	1.6
April	1.2	October	1.9
May	1.3	November	2.5
June	1.6	December	3.4

Present this as a radar chart.



Statistical graphs

Frequency graphs:

When displaying statistical or quantitative data, the most appropriate graphs are a **frequency histogram** and **frequency polygon**.

FREQUENCY HISTOGRAMS:

Are similar to column graphs EXCEPT FOR THE FOLLOWING FEATURES:



Gaps are NEVER left between the columns;



A half unit space is left BEFORE the 1st column;



If the chart is coloured or shaded then it is done all in one colour.



Frequency is ALWAYS plotted on the vertical axis.



For grouped, data, the score is marked so that the CLASS CENTRE of each class appears under the centre of the column.

eg The table below shows the number of people living in each house in a street.

No. of people	Frequency
1	1
2	4
3	10
4	15
5	8

Show this information in a frequency histogram.

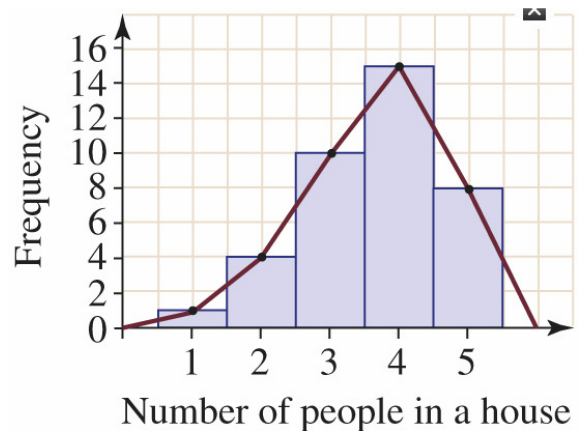


A **frequency polygon** is a line graph that can be drawn by joining the centres of the tops of the columns of the histogram.



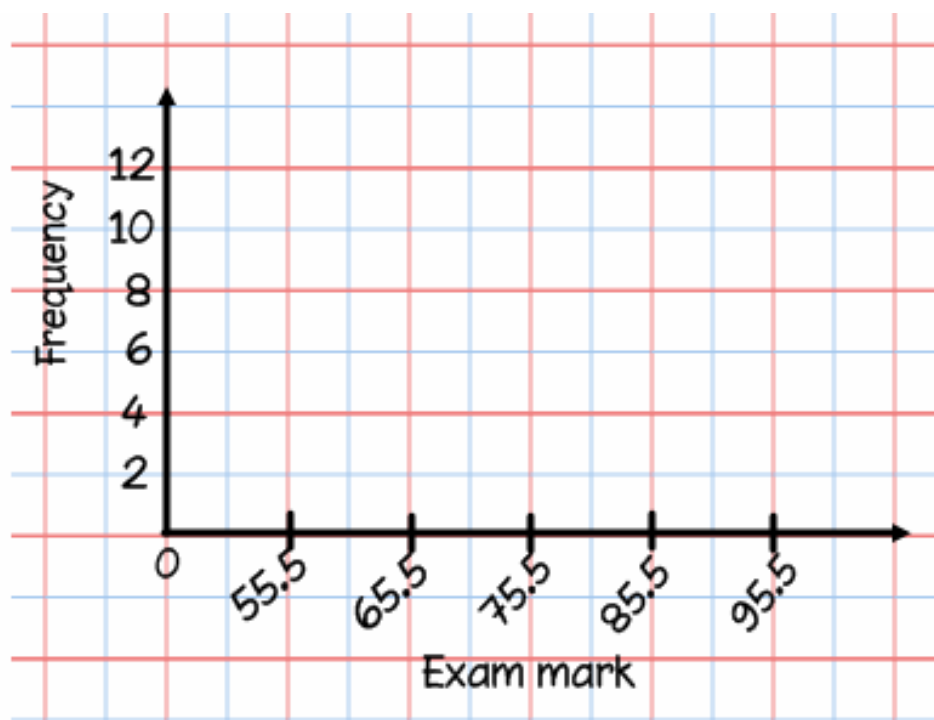
The polygon begins and finished on the horizontal axis a half column width space from the group boundary of the first and last groups.

It is common practice to draw the histogram and the polygon on the same set of axes.



eg The frequency distribution below shows a class set of marks on an exam. Construct a frequency histogram and polygon on the same set of axes.

Mark	Class centre	Frequency
51-60	55.5	3
61-70	65.5	5
71-80	75.5	12
81-90	85.5	7
91-100	95.5	3



Reversing...

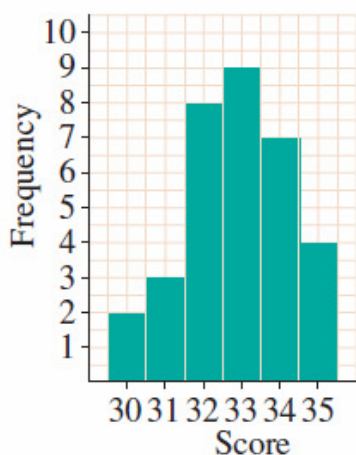
Creating a frequency distribution from a frequency Histogram.

Follow these steps:

- 👤 Construct a table with Score and Frequency;
- 👤 Copy the scores from the horizontal axis into the Score column;
- 👤 The frequency is given by the height of the column above the score.

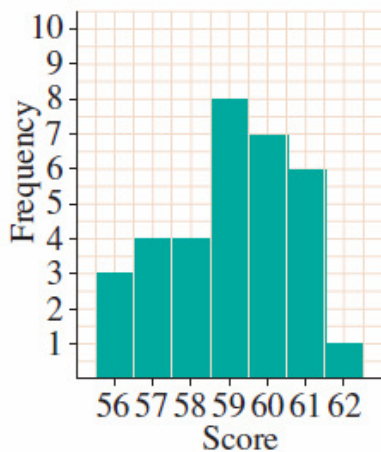
Complete the frequency tables to show the data represented in each frequency histogram.

1



Score	Frequency
30	2
31	
32	
33	
34	
35	4

2



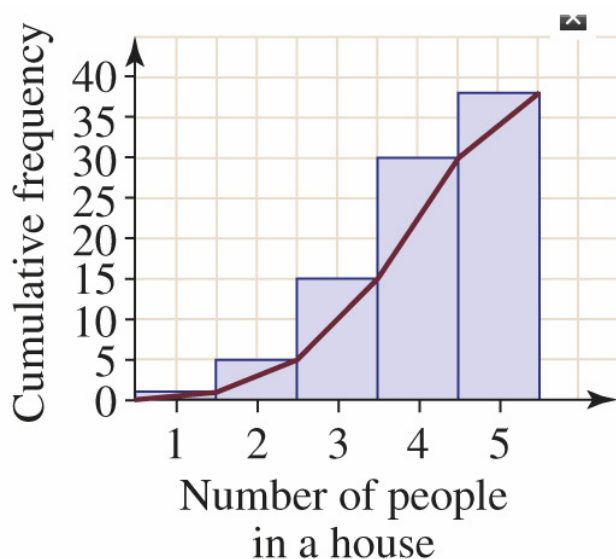
Score	Frequency
56	3
57	4
58	
59	
60	
61	
62	

cumulative graphs

The **cumulative frequency** is a progressive total of the frequency column.

Complete the following table:

No. of people	Frequency	Cumulative frequency
1	1	
2	4	
3	10	
4	15	
5	8	



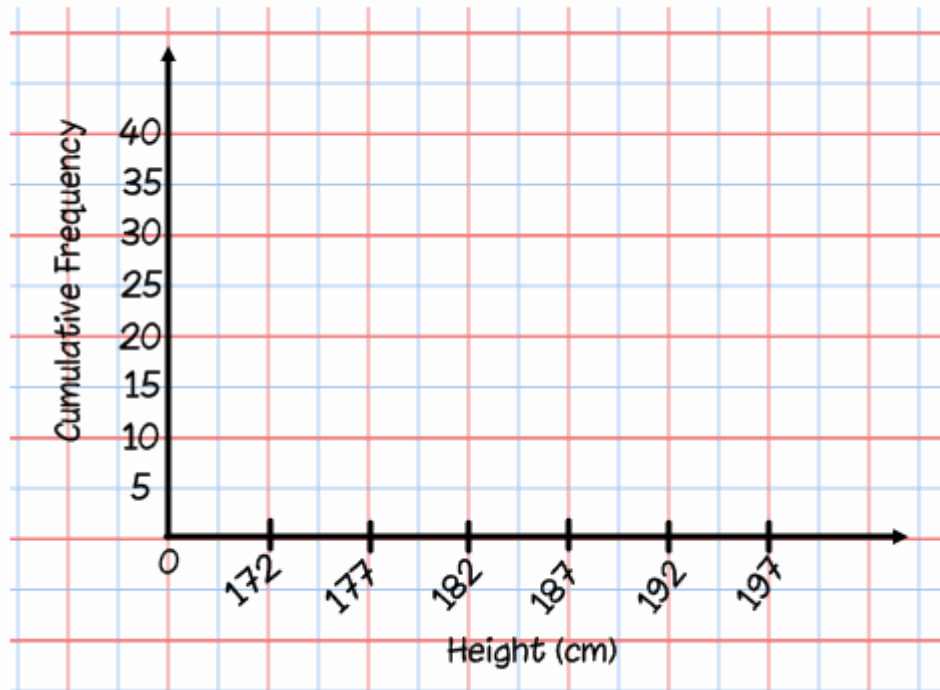
LOOK!

- A CUMULATIVE HISTOGRAM has columns.
- A CUMULATIVE POLYGON (OGIVE) is the line graph.
- Each **MUST NEVER DROP!!!**

eg The frequency table below shows the heights of people in a basketball squad.

Height (cm)	Class centre	Frequency
170–174	172	3
175–179	177	6
180–184	182	12
185–189	187	10
190–194	192	8
195–199	197	1

- Add a cumulative frequency column to the table.
- Construct a frequency histogram and ogive.



Range and Interquartile Range

The range and interquartile range are used to measure the spread of a set of scores.

Range

Range =

eg There are 15 players in the squad for a State Hockey match. The number of State level matches played by each member of the squad is shown below:

2 6 4 12 3 4 0 15 6 3 7 10 7 3 2

What is the range of this distribution? _____

When calculating the range from a frequency distribution, we calculate the range using the score column.

If the data is presented in grouped form, use the highest score from the highest class and lowest score from the lowest class.

eg Find the range of these data:

Height	Frequency
170–175	3
175–180	6
180–185	12
185–190	10
190–195	8
195–200	1


Investigate this!!!

The range however, is often not a good indicator of the overall spread of scores. Consider the two sets of scores showing the marks of two students:

A: 24, 24, 24, 24.5, 25, 25, 26, 80

B: 18, 20, 24, 29, 35, 40, 50, 60

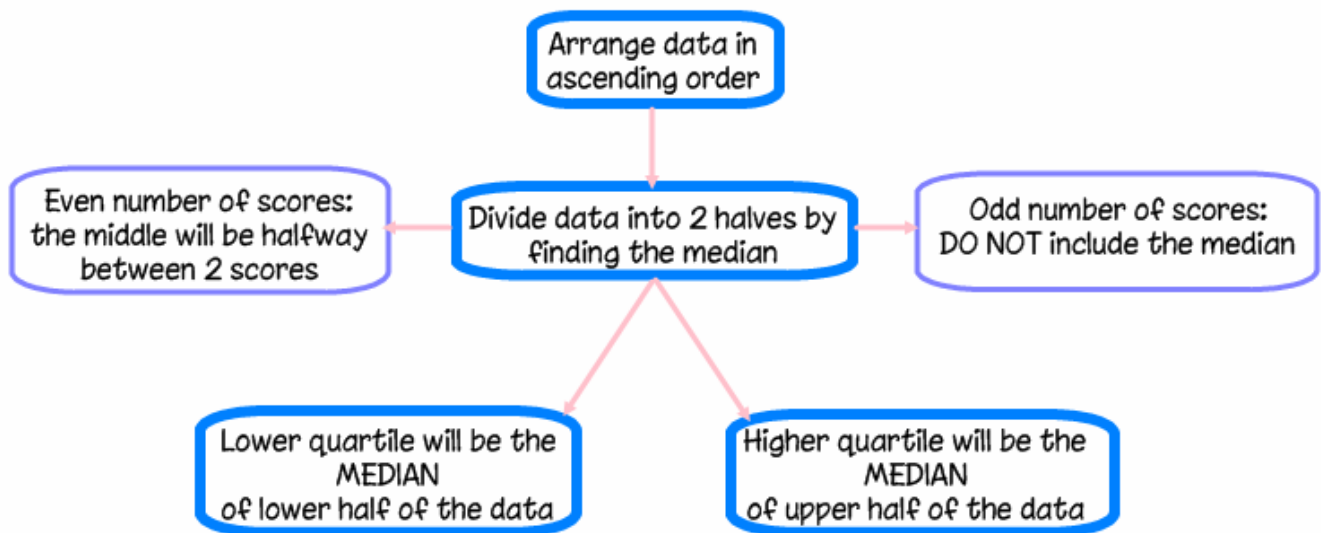
Interquartile Range

- 1 The **quartiles** are found by dividing the data into quarters. The **lower quartile (Q1)** is the lowest 25% of scores; the **upper quartile (Q3)** is the highest 25% of scores.
- 2 The **median** is the middle score (for an odd number of scores) or the average of the two middle scores (for an even number of scores)

eg Calculate the median of:

- a) 2, 5, 8, 8, 8, 11, 12
- b) 34, 64, 23, 57, 89, 44, 62, 76

- 3 The **interquartile range** is the difference of the Upper quartile (Q3) and the lower quartile (Q1). To find the IQR:



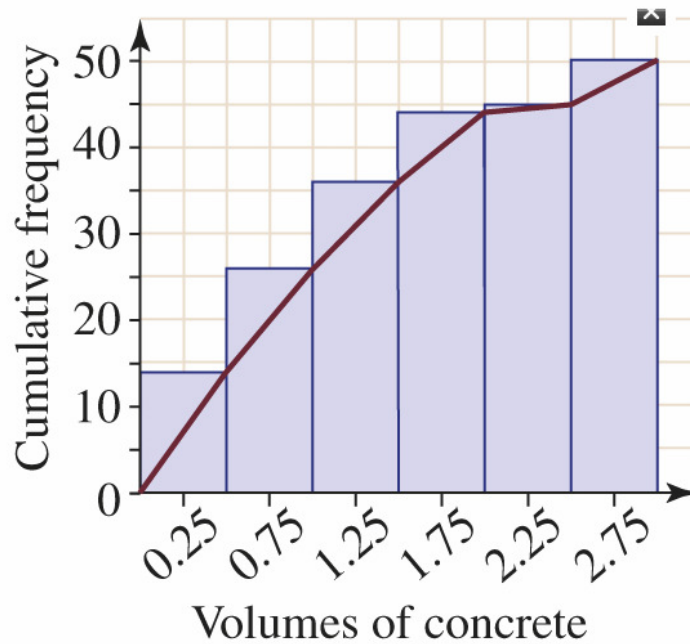
eg Find the interquartile range of the following data:

12, 9, 4, 6, 5, 8, 4, 10, 2

eg The cumulative frequency histogram and polygon at below shows the number of customers who order different volumes of concrete from a readymix concrete company during a day.

Find the:

- a) median;
- b) interquartile range for this distribution



A data set can also be divided into deciles. A **decile** is a band of 10% of all scores, eg the top decile would be found using the 90th percentile.

Stem and Leaf Plots

In a stem and leaf plot, the final digit of each number is shown in the leaf column, while the preceding digit/s is displayed in the stem column.

The numbers in both stem and leaf columns are ordered from **SMALLEST TO LARGEST**. The key on the plot explains how to read the score.

Copy and complete the following stem-and-leaf plot of the given data.
34, 56, 49, 61, 37, 70, 63, 52, 41, 61

THINK

- 1 Copy the given stem-and-leaf plot. The key on the plot tells us that the digit in the tens place of each number makes up a stem and the digit in the units place makes up a leaf.
- 2 Fill in the missing numbers on the plot as follows:
 - i There are two numbers in the set whose stem (that is, the digit in the tens place) is 3. They are 34 and 37. As 34 is already listed, enter 37 by placing its leaf (7) in the space provided.
 - ii There are two numbers in the set whose stem is 4. They are 41 and 49. As 49 is already listed, enter 41 by placing its leaf (1) in the space provided.
 - iii There are two numbers in the set whose stem is 5. They are 52 and 56. Enter these numbers by placing their leaves (2 and 6) in order of increasing size in the spaces provided.
 - iv There are three numbers in the set whose stem is 6. They are 61, 61 and 63. As one of the scores (61) has already been listed, enter the remaining two numbers by placing their leaves in the space provided.
 - v Finally, there is one number in the set whose stem is 7. It is 70. Enter this number by placing its leaf (0) in the space provided.

Key: 3 | 4 = 34

Stem	Leaf
3	4
4 9
5
6	1
7

Find the interquartile range of the data presented in the following stem-and-leaf plot.

Key: 15 | 7 = 157 kg

Stem	Leaf
15	4 8 8
16	1 3 3 6 8
17	0 0 1 4 7 9 9 9
18	1 2 3 3 5 7 8 8 9
19	2 7 8
20	0 2

Try these

For each of the following sets of data complete the stem-and-leaf plots given.

- 1 25, 47, 22, 59, 36, 61, 30, 43, 54, 60

Key: 2|2 = 22

Stem	Leaf
2	2
3
4 7
5
6 1

- 2 70, 66, 85, 90, 93, 84, 69, 71, 88, 67

Key: 6|6 = 66

Stem	Leaf
6	6
7
8 8
9

- 3 101, 122, 117, 135, 104, 112, 139, 140, 143, 126, 129, 137, 136, 125

Key: 10|1 = 101

Stem	Leaf
10	1
11
12 5
13 9
14

- 4 64, 48, 59, 55, 71, 49, 64, 69, 70, 57

Key: 4|8 = 48

Stem	Leaf
4	8
5
6 5
7

- 5 252, 269, 248, 273, 250, 276, 281, 259, 262, 277, 280, 265, 264, 271, 266

Key: 24|8 = 248

Stem	Leaf
24	8
25
26 6
27
28	0

Five-number summaries

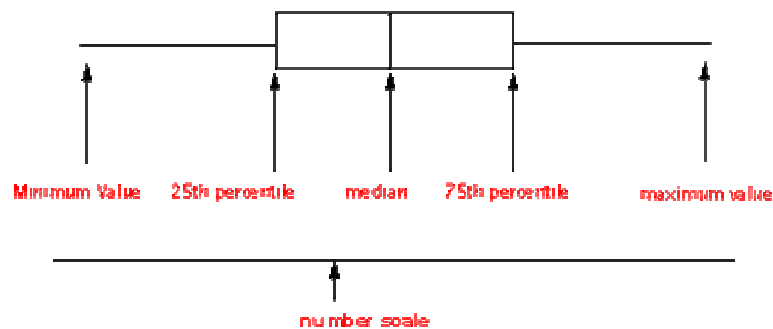
Once we know the median and quartiles, we are able to summarise a data set using 5 numbers. This **5- number summary** consists of:

- 🧐 **minimum score (lower extreme)** - the lowest score in the data set;
- 🧐 **Lower quartile (Q1)** – the score at the 25th percentile
- 🧐 **Median** – the middle score
- 🧐 **Upper Quartile (Q3)** - the score in the 75th percentile
- 🧐 **Maximum score (upper extreme)** – the highest score in the data set

$$\text{Interquartile range} = Q3 - Q1$$

This data is then presented in a **Box & Whisker plot**.

Parts of a Box-and-Whisker Plot



- 🧐 The **box portion** of the box-and-whisker plot **includes 50% of the data**
- 🧐 The **whiskers extend** to the **minimum** and **maximum** data values
- 🧐 **More than one box plot can be drawn** for the number scale allowing comparison of a variable by groups

eg Find the five point summary, Interquartile range and the sixth decile for the following sets of data:

a) 8 10 18 11 7 12 8 9 8 15 12 13 15

Arrange the scores in ascending order:

Minimum value = _____

Q_1 (lower quartile) = _____

Q_2 (median) = _____

Q_3 (upper quartile) = _____

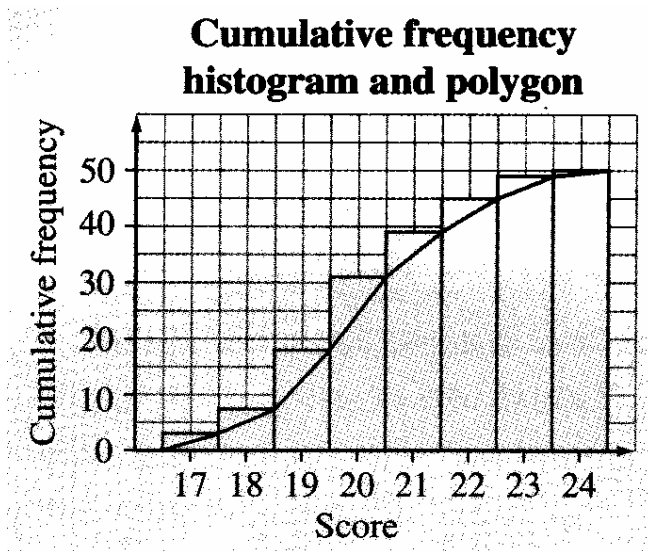
Maximum range = _____

Interquartile range = _____

b)

Stem	Leaf
4	0 1 3
5	2 5 6 7
6	2 8
7	0 3 3 4 7 9
8	3 4 5 6 8
9	0 3 4 5

c)



eg a) Determine the Five-Number Summary for the two data sets of raisin counts displayed below.

b) Construct a box plot for each on the same scale. Based on the box plots, which brand of raisins would you buy? Justify your decision.

Sunripe Raisins

25 25 25 26 26 26 26 26 27 27
27 28 28 28 28 28 28 28 28 28
29 29 29 30 30 31 32 32

Rockin' Raisins

23 24 25 25 25 27 27 27 27 27 27
27 27 28 28 29 29 29 29 29 29 30
31 32 32 33 33 33 34 34 35 35 35
36 36 38

